Full Length Research Paper

Application of bioinformatics to optimization of serum proteome in oral leukoplakia and oral squamous cell carcinoma

Hong He¹, Gang Sun¹, Feiyun Ping²*, Zhijian Xie¹, Huiming Wang⁴*, Yining Li¹ and Kai Zhang¹

¹Stomatology Hospital, School of Medicine, Zhejiang University. 395 Yan an Rd, Hangzhou, 310006, China.
²Second Hospital Affiliated to Medicine College of Zhejiang University 88 Jie fang Rd, Hangzhou 310009, China.

Received 15 November, 2013; Accepted 2 May, 2014

Intervention to the evolution and metastasis of oral squamous cell carcinoma (OSCC) from oral leukoplakia (OLK) out of normal mucosa is currently far from ultimate. New technologies and insights are reckoned on for etiology and clinical protocol of OSCC and mucosa lesions. SELDI-TOF-MS technology, Support vector machine, Discriminate Analysis, and CM10 protein chip were applied to study the sera proteomes of 32 healthy volunteers, 6 patients with oral mucosa leukoplakia, 28 OSCC patients, and 8 patients with metastatic OSCC. Ratios of protein mass to its charge (m/z) showed in valve peak value were delivered and discriminated out of the huge amount of protein data as group markers for identifications. Protein peak values 4181 and 4651 were high in volunteers serum while low in patients with OLK, the sensitivity of which was 100.00% (32/32), specificity was 83.33%(5/6), accuracy was 97.37%(37/38). And m/z 4162, 6886 of 87.82, 92.86 and 66.67%; 4289, 5661, 6195, 4352, 5072 of 97.22, 100.00 and 87.5% were discriminates between OLK and local OSCC, between local OSCC and regionally metastatic OSCC, respectively. Conclusively, researches are encouraged to launch a proteomics assistance and guidance in modern molecular diagnostic approaches for understanding and controlling the mucosa lesions especially in conquering the malignant progress.

Key words: Oral squamous cell carcinoma (OSCC); oral leukoplakia (OLK), ZUCI-PDAS (Zhejiang University Cancer Institute ProteinChip Data Analysis System), bioinformatics technology, discriminate analysis, proteomes of optimization.

INTRODUCTION

Mucosa lesions have never ceased threatening the human's health even with the development of modern science and technology. Oral leukoplakia (OLK) is the first one among the precancerous oral mucosal lesions, with a canceration rate of 7 to 15%, and 14 to 50% in non-homogeneous leukoplakia. Although, studies have

Corresponding authors. E-mail: honghehh@zju.edu.cn. Tel: +86-571-87217431, Fax: +86-571-87217423.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abbreviations: Oral squamous cell carcinoma (OSCC), oral leukoplakia (OLK)
illuminated that there exist in biomarkers as cytokines, genes, and antigens in the precancerous lesions to escape the failure in interfering the marching of malignance and metastasis of tumor. The reality is that 80% of oral malignant tumors, which rank the sixth in the queue of human malignant tumors, are the oral squamous cell carcinoma (OSCC), whilst the deformation and recurrence rate after surgery treatment are high and the prognosis is out of anticipation (Chen et al., 2004; Ferlito et al., 2002). To search for any meritorious biomarkers to trace the lesions evolution, scientists found that in the mass spectrometer, samples could be respired by laser irradiation and dissociated into gasified ions, in the uniform electric field conditions these ions perform an accelerative out fly, fly through a vacuum tube with no electric field, and eventually captured by the ion detecting receiver. The square of the ion flight time is in inverse proportion to the ratio of molecule size and molecular charge, when the numbers of charge (ions evoked by Surface Enhanced Laser Desorption/Ionization SELDI (He et al., 2009; He et al., 2011; Hu et al., 2005) are usually a single charge are equal, the time of flight associates with the molecule size, the smaller the molecule is, the shorter the flight time would be, hence small molecules were first to reach the receiver. If reflected in a mass spectrum diagram, the peak position corresponds to molecule weight of the relevant protein compositions; and the spectral peak height corresponds to the number of the relevant protein compositions. Thereby proteins could be identified with the peak position and height. This is what the protein fingerprint is. This research is aimed to obtain a sequence of fingerprints of serum protein between OLK and OSCC.

MATERIALS AND METHODS

Serum sample collection and processing

Under informed consent, sera samples of patients and healthy volunteers from the Affiliated Stomatology Hospital and 2nd Hospital of Zhejiang University during Feb 2010 to March 2014 were obtained before any treatment was implemented, and collected in the early morning before breakfast, then immediately separated and stored at -80°C until use, with diagnoses confirmed by post surgical pathology. There were 19 males and 9 females in the 28 cases of local OSCC (L. OSCC) group with a median age as 56.6 years (35 to 68 years range). Four males and 2 females with a median age as 53 years (44 to 59 years range) constituted the OLK group. Seven males and 1 female with a median age as 56.4 years (37 to 71 years range) constituted the regional metastasis group (R. OSCC). 22 males and 1 female with a median age as 50.9 years (34 to 71 years range) constituted the volunteers group (N.).

Proteins bind chip and spectrometer detection

After thawing and 2 min of centrifugation (10,000 r/min), 5 μl serum sample was added into 10 μl 0.5% U9 (9 mol/L urea, 0.2% CHAPS (3[3-cholamidopropyl] dimethylammonio)-1-propane sulfonate), 0.1% DTT (DL-dithiothreitol) in a 96-well plate and incubated for 30 min at 4°C with 600 r/min vigorous shaking. Supernatant was collected and the procedure is repeated once. Then, 185 μl NaAc was added into each well in the 96-well plate (600 r/min, 2 min) and 100 μl samples of different patients disposed as above were separately added into different well of the ProteinChip array cassette (600 r/min, 1 h). After the content from each well was removed, each well was washed with 200 μl NaAc (pH = 4.0, 600 r/min, 5 min). This procedure was repeated two more times. Each spot was washed with 200 μl HPLC water, which was removed immediately. This was repeated once. After natural air drying, 1 μl SPA (sinapic acid) was applied to each spot. After natural drying for another 5 min, another 1 μl SPA was applied. Naturally dried again.

American Ciphergen SELDI Protein Biology System II plus (PBS II plus) and ProteinChip Software (Version 3.0, Ciphergen Biosystems) were used to read the chips and analyze the data. Setting parameters: laser intensity 165, 65 laser shots per sample, detector sensitivity 7, automatically detected peaks from 2,000 to 30,000 m/z. Mass accuracy was calibrated to less than 0.1% using the All-in-1 peptide molecular mass standard chip (Ciphergen Biosystems). The peaks were normalized and noises were filtrated (first Signal to Noise ratio >2.5). Peak clusters were completed using second-pass peak selection (Signal to Noise ratio>4, within 0.3% mass window) and estimated peaks were added. Discreant mass peaks of different groups were identified by support vector machines (SVM) applying radially basis kernel function with Gamma value as 0.6, Penalty score function C as 19, each m/z peak was proved with Wilconxon rank test (p < 0.05).

RESULTS

After the tremendous comparison data from the mass spectrums, 112, 108, 120 discrepant protein mass peaks between sera of 32 healthy volunteers and 6 OLK patients,
28 OSCC patients, 28 locally sited OSCC patients and 8 regionally metastatic OSCC were respectively exhibited as significant difference, among which ten of the lowest $p$ value of the integrated proteomes constituted respectively of 2, 2, 5 protein peaks (Figure 1a to c, Tables 1 to 4) were eventually selected out as the highest Youden’s index via SVM Discriminate Analysis and as common key sense upon corresponding and simultaneous changes. They were 4181 mutually together with 4651; 4162 together with 6886; 6195, 5661, 4289~~ and 4352 together with 5072. These ultimately obtained individual protein peaks and integrated discrepant serum proteomic selections with sensitivity, specificity, and accuracy and with some of the peak figures were presented as shown in Figures 1a to c.

DISCUSSION

The expression profiling of serum proteins from human with OLK or OSCC and healthy volunteers was investigated by a combination of proteomic techniques in this study. Different to Panicker’s research (Kordy et al., 2012), which is currently lack of diseases links such as to carcinoma with metastasis or not, and in some cases to pre-invasive disease, and although our study is temporally lack of protein identification, our present research can be used to further some understanding of the serum proteomic behaving from individual to universal of oral mucosal lesion for tumorgenesis etiology as well as tumor evolution, and provide a snapshot of the proteome for diagnosis, prognosis and prophylaxis of oral mucosal disease. Validated protein biomarkers could be useful in early detection of disease and any possibility to be getting malignant or metastatic, monitoring disease progression or monitoring response to treatment. Upon neoplasia arises, it is likely to study proteins produced by the local lesion as well as by the host reaction in serum with response to the lesion. Advances in proteomic technologies have greatly accelerated the field of protein biomarker discovery (Nemet et al., 2005; Panicker et al., 2009; 2010). High-throughput technology Surface Enhanced Laser Desorption and Ionization Time of Flight Mass Spectrometry (SELDI-TOF-MS) has achieved tremendous harvests from different cancerous field to infectious diseases or drugs for researches (Sun et al., 2012).
Figure 1b. Serum protein mass/charge wave peak 6886 expressed statistically different between the group of patients with OLK (group 0) and the group of patients with OSCC (group 1), $P=0.0015641816$, together with m/z 4162 protein, established the distinguished proteome with a highest Youden’s index.

Figure 1c. Serum protein mass/charge wave peak 5072 expressed significantly low in the group of patients with metastasis OSCC (group 0), while remarkably high in group of patients with local OSCC tumor (group 1), although $P=0.1529314178$, it was one of the distinguished proteome with a highest Youden’s index between the two compared groups.
Table 1. Discrepant protein peaks between OLK patients and healthy volunteers (x±SD).

<table>
<thead>
<tr>
<th>m/z Peak</th>
<th>p value</th>
<th>Healthy volunteer (group 0)</th>
<th>OLK patient (group 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4651</td>
<td>0.00000086626</td>
<td>1905.30±548.37</td>
<td>686.69±386.12</td>
</tr>
<tr>
<td>4181</td>
<td>0.0001307483</td>
<td>5901.24±2548.82</td>
<td>1333.55±1042.86</td>
</tr>
</tbody>
</table>

Table 2. Discrepant protein peaks between OLK patients and local OSCC patients (x±SD).

<table>
<thead>
<tr>
<th>m/z Peak</th>
<th>p value</th>
<th>OLK patient (0)</th>
<th>Local OSCC patient (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6886</td>
<td>0.0015641816</td>
<td>3551.45±1833.51</td>
<td>1818.32±921.62</td>
</tr>
<tr>
<td>4162</td>
<td>0.0365123759</td>
<td>1101.78±373.05</td>
<td>1644.09±579.45</td>
</tr>
</tbody>
</table>

Table 3. Discrepant protein peaks between local OSCC patients and regional metastatic OSCC patients (x±SD).

<table>
<thead>
<tr>
<th>m/z Peak</th>
<th>p value</th>
<th>R. metastatic OSCC (1)</th>
<th>Local OSCC patient (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6195</td>
<td>0.0015387984</td>
<td>893.32±382.21</td>
<td>578.99±165.48</td>
</tr>
<tr>
<td>5661</td>
<td>0.0061131375</td>
<td>771.95±255.40</td>
<td>1167.44±355.59</td>
</tr>
<tr>
<td>4288</td>
<td>0.0289773483</td>
<td>1865.87±772.37</td>
<td>2891.56±1196.07</td>
</tr>
<tr>
<td>4352</td>
<td>0.0540207442</td>
<td>3663.07±946.03</td>
<td>2708.56±1249.09</td>
</tr>
<tr>
<td>5072</td>
<td>0.1529314178</td>
<td>1010.51±384.44</td>
<td>1518.43±952.59</td>
</tr>
</tbody>
</table>

Table 4. The sensitivities, specificities and accuracies of integrated discrepant serum proteomes in OLK, local (L) and regional (R) metastatic OSCC patients, and healthy volunteers.

<table>
<thead>
<tr>
<th>Sample (0 vs 1)</th>
<th>m/z Peaks proteome</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N. vs OLK</td>
<td>4181, 4651</td>
<td>97.37</td>
<td>83.33</td>
<td>100.00</td>
</tr>
<tr>
<td>OLK vs L. OSCC</td>
<td>4162, 6886</td>
<td>87.82</td>
<td>92.86</td>
<td>66.67</td>
</tr>
<tr>
<td>L. vs R. OSCC</td>
<td>4289, 5661, 6195, 4352, 5072</td>
<td>97.22</td>
<td>100.00</td>
<td>87.5</td>
</tr>
</tbody>
</table>

Note: Italic values of protein peak represent high expressions in group 0; regular scripts represent high expression in group 1.

2009; Weinberger et al., 2000; Xu et al., 2006; Yu et al., 2004; Yu et al., 2005; Zhou et al., 2013). Of the current proteomic tools, no single method can resolve an entire proteome. Combination of several methods like Zhejiang University Cancer Institute ProteinChip Data Analysis System (ZUCI-PDAS) with SELDI-TOF-MS (He et al., 2009; 2011; Hu et al., 2005) and SVM established the Bioinformatics. One of the merits of this study protocol is the Discriminate Analysis derived from ZUCI-PDAS and SVM has launched the other merit of it to link to the development of diseases for the biological meaning of the different preferential expressions (He et al., 2009; 2011; Hu et al., 2005).

On the one hand, SVM technique is proper to solve the limitation of samples, and on the other hand, our future work still would take into consideration to enlarge the samples as well as about the necessity to identify the proteins in the optimized group peaks. Current study indicates that progressive study of such prognostic biomarkers which were based on tumor phenotype and biologic behavior, no matter proteomes of protein mass peaks or characterization of individual protein, would allow clinicians not only to diagnose a disease involving OSCC as well as precancerous lesion like OLK, but also to select the most efficient treatment modalities other than absolutely radical surgery.

ACKNOWLEDGEMENTS

This work is funded by the Projects and funds sponsored by Scientific Research Foundation for the Returned Overseas Chinese Scholars, HR Office of Zhejiang Province (J20120036), Qianjiang Talent Scheme.
Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES


